



## DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets <sup>7</sup> : <b>H04N 7/15, 7/14</b>		<b>A1</b>	(11) Numéro de publication internationale: <b>WO 00/13417</b>
			(43) Date de publication internationale: 9 mars 2000 (09.03.00)
<p>(21) Numéro de la demande internationale: PCT/FR99/02047</p> <p>(22) Date de dépôt international: 26 août 1999 (26.08.99)</p> <p>(30) Données relatives à la priorité: 98/10888 31 août 1998 (31.08.98) FR</p> <p>(71) Déposant: FRANCE TELECOM [FR/FR]; 6, place d'Alleray, F-75015 Paris (FR).</p> <p>(72) Inventeurs: VIALLET, Jean-Emmanuel; 12, chemin de Toul Al Lann, F-22370 Tregastel (FR). FERAUD, Raphaël; 7, rue du Machallac'h, F-22300 Lannion (FR). COLLOBERT, Michel; 4, Gwéradur Vraz, F-22560 Pleumeur Bodou (FR). BERNIER, Olivier; 50A, rue Kervenno, F-22300 Lannion (FR).</p> <p>(74) Mandataire: BORIN, Lydie; Cabinet Ballot-Schmit, 16, avenue du Pont Royal, F-94230 Cachan (FR).</p>			<p>(81) Etats désignés: CN, JP, brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p><b>Publiée</b> <i>Avec rapport de recherche internationale.</i></p>

**(54) Title:** AUTOMATIC SYSTEM FOR SOUND AND IMAGE RECORDING

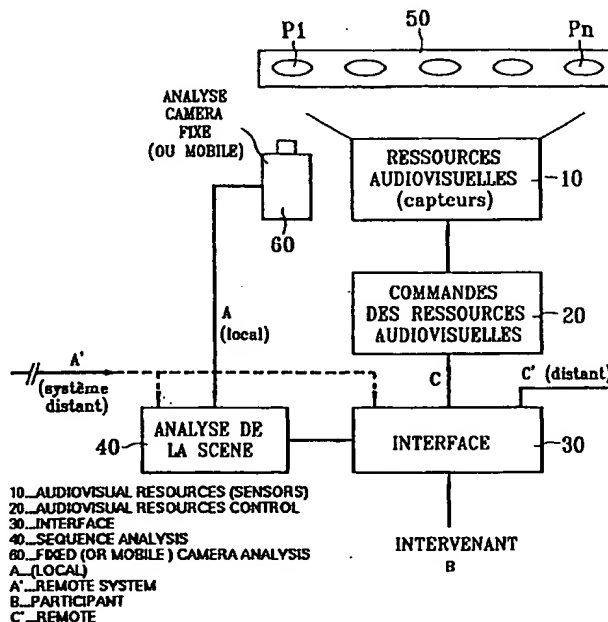
**(54) Titre: SYSTEME AUTOMATIQUE DE PRISE DE SON ET D'IMAGES**

**(57) Abstract**

The invention concerns an automatic system for sound and image recording in particular for videoconference, comprising means controlling (20) image and sound recording sensors (10) and sequence analysing means (40) monitoring said control means (20) to obtain automatic framing of the sequence being filmed. The invention is characterised in that an intelligent interface is provided for selecting a person or a group among the persons in the filmed sequence, at the request of a participant, and automatic framing (30) on the basis of data supplied by the sequence analysing means (40), on the selected person or group. The invention is applicable to videoconferencing.

**(57) Abrégé**

L'invention concerne un système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande (20) de capteurs de prise de vues et de son (10) et des moyens d'analyse de scène (40) pilotant ces moyens de commande (20) pour obtenir un cadrage automatique de la scène filmée. Selon l'invention il est prévu une interface intelligente pour réaliser la sélection (30) d'une personne ou d'un groupe parmi des personnes de la scène filmée, sur commande d'un intervenant, et le cadrage automatique (30) à partir des informations fournies par les moyens d'analyse de la scène (40), sur la personne sélectionnée ou le groupe. Application à la visioconférence.



# UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

## SYSTEME AUTOMATIQUE DE PRISE DE SON ET D'IMAGES

L'invention concerne un système automatique de prise de son et d'images notamment pour la visioconférence.

Le besoin de tels systèmes se fait sentir surtout  
5 dans le domaine de la visioconférence, mais peut également être intéressant dans d'autres situations mettant en oeuvre de la prise de vue avec plusieurs personnes par exemple.

Il s'agit donc de systèmes avec lesquels les prises  
10 de son et de vue ne sont pas pilotées par une ou plusieurs personnes (cameraman et ingénieurs du son dans le domaine de la télévision) mais par un équipement automatique.

Dans le cas de la visioconférence, un intervenant a  
15 cependant la possibilité avec certains systèmes équipés d'une télécommande de piloter la prise de vue d'une caméra, mais ces systèmes présentent des inconvénients comme on va le voir dans la suite.

En effet, des systèmes de visioconférence sont  
20 équipés de moyens de prise de vues et de son, disposant d'équipements (caméras et microphones) qui ne sont pas orientables ou dont l'orientation est commandée par l'intermédiaire d'une télécommande.

Dans des systèmes de visioconférence existants,  
25 diverses caméras télécommandables et contrôlables à distance via une communication série sont utilisés.

La télécommande permet de balayer continûment le site et l'azimut de la caméra ainsi que de faire varier continûment le zoom de la caméra. L'orientation de la  
30 caméra dans la direction occupée par une personne ou un groupe de personnes est possible, mais malaisée.

Des directions de l'espace (aux nombres de six pour les deux caméras ) peuvent être mémorisées par la caméra. La caméra peut être dirigée dans l'une de ces directions par activation d'une touche de la télécommande ou par contrôle du port série. L'intérêt de cette fonction est d'accéder directement à une direction de l'espace sans avoir à agir par combinaison de touches successives (site, azimuth). Ainsi, lorsque les positions de l'espace préenregistrées correspondent à des places autour d'une table, l'utilisateur de la télécommande peut simplement passer d'une personne à une autre.

Il existe également des systèmes reposant sur la localisation acoustique selon lesquels l'orientation d'une caméra est pilotée à partir de l'analyse acoustique de la scène.

L'analyse acoustique de la scène est obtenue à partir de plusieurs microphones qui permettent de déterminer la direction des sources sonores, voire des sources de parole. La direction des sources de parole étant identifiée, elles pourraient être sélectionnées une à une, puis être suivies dynamiquement. La fonction Lime Light de Picture Tel, entreprise qui réalise et commercialise des systèmes de visioconférence, est basée sur la localisation acoustique et permet la détection et le suivi d'une source sonore et l'orientation dynamique d'une caméra.

Il existe également des systèmes reposant sur la localisation visuelle de personnes. Cette localisation visuelle est basée sur l'analyse par traitement d'image issue de la caméra.

Si le champ de la caméra est suffisamment large, et le système performant, toutes les personnes présentes seront détectées. A la différence de la localisation

acoustique, la localisation visuelle bénéficie de la permanence des visages à l'image et de l'absence de réflexion (absence de miroirs dans les salles de visioconférence et dans les lieux habituels de prise de vues). Plusieurs techniques peuvent être utilisées  
5 séparément ou non pour détecter des personnes : détection de mouvement, détection de la teinte de la peau, détection de formes ovales, détection globale de visages (par réseau de neurones par exemple), détection  
10 d'attributs de visages tels les yeux, la bouche.

Un certain nombre de systèmes de détection de visages existe et certains associent une fonction de prise de vue automatique. On peut citer à titre d'exemples les systèmes qui sont décrits dans les  
15 publications suivantes :

P1 - M. Collobert, R. Feraud, G. Le Tourneur, O. Bernier, J.E. Viallet, Y. Mahieux, D. Collobert, "LISTEN : a System for Locating and Tracking Individual Speakers", 2nd Int. Conf. On Automatic Face and Gesture  
20 Recognition, Killington, USA, October 1996, pp 283-288.

P2- Hunke, Locating and Tracking human faces with neural network, Technical Report CMU-CS-94-155, 1994.

P3 - Yang, Wu, Waibel, Focus of attention in video conferencing, Technical Report CMU-CS-96-150, 1996.

25 Cependant, parmi tous les systèmes qui viennent d'être décrits, un grand nombre présente des inconvénients ou de toute façon ne sont pas satisfaisants pour les besoins qui se font sentir actuellement et qui sont précisés dans la suite.

30 - En effet, les systèmes ayant une fonction de balayage de site, azimuth et focale au moyen d'une télécommande présentent les inconvénients suivants :

Pour obtenir un cadre adapté, l'utilisateur doit appuyer successivement sur plusieurs touches de la

télécommande (site, azimut et zoom) avant d'obtenir le résultat désiré. Cette fonctionnalité est peu pratique, lente, sujette à des erreurs de manipulation et est donc peu utilisée. Elle est de fait impraticable si l'on souhaite adapter périodiquement ou continûment le cadre de la caméra à une personne qui se déplace ou un ensemble de participants qui naturellement bougent. De la même manière, cette fonctionnalité est également impraticable quand on souhaite changer régulièrement de personne.

- La fonction accès par la télécommande à des directions préenregistrées de l'espace présente les inconvénients suivants :

Le premier inconvénient est lié au fait que les positions doivent être préenregistrées. Elles ne peuvent donc être rapidement continûment modifiées.

Le second inconvénient est de supposer que les personnes occuperont bien les positions préenregistrées et qu'elles n'en bougeront pas. Dans la pratique et même en fixant les chaises au sol, on constate que les personnes bougent et donc sont rarement au centre du cadre, voire sortent du cadre si celui-ci est serré sur la personne. Cet inconvénient est manifeste dans le cadre du système de visioconférence où spontanément les personnes sortent du cadre défini par les directions préenregistrées de l'espace.

L'inconvénient suivant est que cette fonctionnalité d'orientation de la caméra dans une direction de l'espace n'associe pas toujours de focale spécifique. Ainsi, l'utilisateur de la télécommande doit actionner les touches de contrôle de la télécommande pour adapter la taille du cadre à la position effectivement occupée par une personne. En effet, si la personne a tendance à sortir du cadre, en passant à une focale plus courte,

cette personne revient dans le cadre sans avoir besoin de changer la direction de l'espace dans laquelle pointe la caméra. Mais alors le cadre n'est plus serré sur la personne et, il y a éventuellement plusieurs  
5 personnes dans le cadre.

Lorsqu'il est possible d'associer et de mémoriser une focale à chacune des directions mémorisées de l'espace, cela implique que d'une part, l'utilisateur détermine au préalable la focale qu'il estime adaptée  
10 et d'autre part, la mémorise. D'autre part, cette focale enregistrée est fixe et donc non adaptative et, comme pour la direction, plus ou moins simplement modifiable.

La fonctionnalité d'accès à des directions  
15 prédéterminées de l'espace peut convenir à certaines situations stables (télésurveillance), mais ne permet pas de s'adapter à une situation particulière. La caméra pointe dans la direction de l'espace, mais ne sait rien du contenu de l'espace, qu'il soit occupé par  
20 une personne ou vide.

Un autre inconvénient, secondaire, est le nombre limité à 6 directions de l'espace qui peuvent être mémorisées par la caméra et donc accessibles par la télécommande. Cet inconvénient est en général résolu en  
25 mémorisant ces directions dans un ordinateur et en utilisant une télécommande avec un plus grand nombre de touches.

La localisation acoustique présente les inconvénients suivants :

30 En raison de la présence de sources sonores parasites (bruits de chaises, de portes, d'appareils, ...) superposées à la parole des personnes présentes et en raison de réflexions multiples des

sources acoustiques sur les murs, la localisation acoustique est souvent peu précise.

D'autre part, l'activité acoustique de parole est par nature intermittente (lorsqu'une personne s'arrête de parler pour écouter).

Enfin, la localisation acoustique est sensible à l'amplitude de la source sonore.

La localisation visuelle présente les inconvénients suivants :

Le principal inconvénient de la localisation visuelle est lié à la complexité des algorithmes, à leur rapidité et à leur robustesse. Néanmoins, plusieurs systèmes sont opérationnels soit sur station de travail soit ordinateur personnel (PC) comme les systèmes développés par le déposant, ou comme dans les publications citées précédemment par le déposant.

La fonction de prise de vue automatique d'un groupe de personnes réalisée par le déposant est, à l'usage, particulièrement utile bien que complexe. Le cadre s'adapte en permanence au nombre et la position des participants d'une visioconférence.

Cependant, il apparaît à l'usage, que bien que plus performant et plus satisfaisant que tous les autres, il soit insuffisant car, comme cela a été dit, des besoins se font sentir sur les possibilités d'effectuer à la demande un cadrage automatique sur l'un ou l'autre des participants et de pouvoir revenir tout simplement au cadre général.

L'invention propose pour cela une interface intelligente apte à réaliser la sélection d'une personne (ou d'un groupe de personnes) parmi les personnes de la scène filmée, sur la commande d'un intervenant, et le cadrage automatique à partir des



informations fournies par l'analyse de la scène, sur la personne sélectionnée (ou le groupe de personne).

L'invention a donc pour objet un système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande de capteurs de prise de vues et de son et des moyens d'analyse de scène pilotant ces moyens de commande pour obtenir un cadrage automatique de la scène filmée.

Le système comporte des moyens de sélection d'une personne ou d'un groupe de personnes parmi les personnes de la scène filmée et des moyens de cadrage automatique à partir des informations fournies par les moyens d'analyse de la scène, sur la personne sélectionnée ou le groupe de personne.

L'invention a plus particulièrement pour objet, un système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande des capteurs de prise de vues et de son , des moyens d'analyse de scène pour fournir des signaux de position aux moyens de commande, des moyens de sélection d'une personne ou d'un groupe parmi des personnes de la scène filmée,

-principalement caractérisé en ce que les moyens de sélection comportent une interface physique comprenant une commande à distance apte à permettre la sélection de l'une quelconque des personnes de la scène ou d'un groupe, pour avoir un cadrage automatique autour de cette personne ou du groupe, ou de sélectionner l'ensemble des personnes pour avoir un cadrage général de la scène ;

- et en ce que les moyens de cadrage comportent une interface logique apte à établir une correspondance entre la personne sélectionnée par la commande à

distance et les informations de position issues de l'analyse de scène pour fournir aux moyens de commande les informations de position de cette personne ou du groupe par rapport à la scène filmée.

5        Selon un mode de réalisation de l'invention, la commande à distance est une télécommande universelle, activant un dispositif apte à émettre des signaux de commande à l'interface logique .

10       Les signaux émis par la télécommande peuvent être infra rouge ou électromagnétiques.

      Les signaux de commande de ladite télécommande peuvent être reçus et ré-émis par un émetteur-récepteur.

15       Les signaux de commande de ladite télécommande peuvent être reçus et ré-émis par un dispositif de reconnaissance de la parole ou de reconnaissance de geste

20       La commande à distance peut être réalisée par la télécommande de la caméra d'analyse d'image, les signaux de commande de ladite télécommande étant reçus et ré-émis par la caméra d'analyse vers l'interface logique.

25       Selon un autre mode de réalisation, la commande à distance est une télécommande universelle, les signaux de commande de ladite télécommande étant reçus et ré-émis par la caméra d'analyse.

      Selon un autre mode de réalisation, la commande à distance comprend une interface graphique.

30       La commande à distance comporte en outre , dans ce cas, un écran sur lequel sont visualisées la scène et les différentes zones sélectionnables.

      La commande à distance comporte un périphérique d'entrée/sortie d'ordinateur pour sélectionner les zones repérées.

Dans un mode de fonctionnement, il peut être prévu que les moyens d'analyse de scène reçoivent un signal d'analyse local (A), que les moyens de sélection  
5 permettent alors de sélectionner une personne ou un groupe de personnes de la scène filmée localement et que les moyens de cadrage automatique utilisent les informations de la scène filmée localement.

Dans un autre mode de fonctionnement, il peut être  
10 prévu que les moyens d'analyse reçoivent un signal (A') d'un système distant pour ou correspondant à l'analyse de scène et que les moyens de sélection permettent alors de sélectionner une personne ou un groupe de personnes de la scène filmée à distance et les moyens  
15 de cadrage automatique permettent de piloter le cadrage de la scène filmée à distance, les signaux de commande étant transportés vers le système distant.

D'autres avantages et particularités de l'invention  
20 apparaîtront à la lecture de la description qui est donnée à titre indicatif et nullement limitatif et en regard des dessins annexés sur lesquels :

- la figure 1, représente un schéma de principe de l'invention,
- 25 - la figure 2, représente un schéma plus détaillé de l'invention,
- la figure 3, représente un mode de réalisation particulier pour l'interface physique,
- la figure 4, représente un autre mode de  
30 réalisation pour l'interface physique,
- la figure 5, représente un autre mode de réalisation de l'interface physique,
- la figure 6, représente un autre mode de réalisation de l'interface physique,

- la figure 7, représente un autre mode de réalisation de l'interface physique.

Sur la figure 1 on a représenté de façon  
5 schématique un système automatique de prise de son et  
d'images dans lequel on dispose de ressources  
audiovisuelles 10 pour filmer et capter le son d'une  
scène 50.

Dans le cas d'un système de visioconférence la  
10 scène est composée d'une ou plusieurs personnes dites  
intervenants P1-Pn sur un site, désirant communiquer  
avec d'autres personnes d'un site distant.

Les ressources audiovisuelles 10 sont constituées  
par des capteurs audio et visuels.

15 Les capteurs audio sont par exemple une série de  
microphones placés à proximité des intervenants. Les  
capteurs vidéo sont constitués par une ou plusieurs  
caméras filmant la scène.

Les ressources audiovisuelles 10 sont pilotées par  
20 un dispositif de commande 20 classique, apte à fournir  
les signaux de commande aux capteurs 10 selon les  
informations reçues en entrée par l'interface 30 comme  
cela est détaillé dans la suite.

Conformément à l'invention, les informations reçues  
25 en entrée sont fournies par l'interface 30 à partir du  
dispositif d'analyse de scène 40 et de la sélection  
effectuée par un intervenant.

Le dispositif d'analyse de scène peut être soit  
audio, soit visuel, soit audiovisuel associé à des  
30 capteurs visuels ou audiovisuels.

Dans l'exemple de réalisation qui est donné, ce  
dispositif est un dispositif visuel existant. A cette  
fin, on utilise une caméra d'analyse fixe 60, (la  
caméra peut être mobile), qui permet de fournir le

signal requis servant à effectuer une analyse de la scène visuelle observée.

Le dispositif d'analyse de la scène comprend par conséquent à cet effet, la caméra 60 et des moyens de traitement 40 du signal A fourni par cette caméra. Ces  
5       moyens sont réalisés par exemple par un micro-ordinateur ou une station de travail équipée d'un programme spécifique, existant, d'analyse de scène.

De façon plus précise selon l'exemple de  
10       réalisation mis en pratique, les visages des personnes présentes dans le champ visuel, sont détectés par un réseau de neurones, puis ledit programme met en oeuvre un algorithme qui suit les visages détectés. D'autres techniques connues peuvent être utilisées.

15       Selon un autre mode de réalisation, on peut utiliser un dispositif d'analyse de la scène 40 avec une caméra mobile.

      Selon un troisième mode de réalisation, on peut utiliser ou réaliser un dispositif d'analyse de la  
20       scène utilisant plusieurs caméras fixes ou mobiles.

      Les différents capteurs 10 sont commandés par un dispositif de commande 20 qui reçoit des signaux de commande de l'interface 30 conformément à la présente invention.

25       Dans le mode de réalisation qui a été réalisé, illustré par le schéma de la figure 2, il s'agit d'un dispositif de commande 20 d'une caméra motorisée 11 qui assure la prise de vue et d'une antenne acoustique 12 qui assure la prise de son.

30       On va détailler ci-dessous, la prise de vue et de son pour un ensemble de personnes et pour une seule personne, ce qui correspond à des réalisations effectives pour le déposant.

Les mêmes techniques peuvent être mises en oeuvre pour une prise de vue et de son concernant un groupe de personnes ; le groupe est un sous-ensemble de l'ensemble des personnes.

5 Dans le mode de réalisation préféré qui a été réalisé, l'analyse de la scène est visuelle, c'est-à-dire que l'on détermine la position des personnes mais on ignore si elles parlent.

10 Dans un autre mode de réalisation impliquant une analyse audiovisuelle (incluant donc le son ) les dispositifs de prise de son seront sélectionnés à partir des informations audiovisuelles.

15 Lorsque l'utilisateur du système demande, à travers l'interface 30, une prise de vue sur l'ensemble de personnes P1-Pn, le dispositif de commande 20 commande la caméra 11 de manière à ce que l'ensemble des personnes, présentes dans le champ d'analyse soient cadrées, en respectant les règles de l'art de la prise de vue dans la mesure où les contraintes de la caméra 11 l'autorisent.

20 Lorsque l'utilisateur du système demande, à travers l'interface 30, une prise de vue sur une personne en particulier, le dispositif 20 commande la caméra 11 de manière à ce que la personne, en respect avec les règles de la prise de vue, soit latéralement centrée, que ses yeux soient approximativement au tiers supérieur de l'image par exemple.

30 Dans un mode de réalisation mis en oeuvre, la personne sélectionnée étant encadrée par un ou plusieurs voisins, la prise de vue cherche à isoler cette personne des autres à l'image, dans la mesure où les contraintes liées à la caméra et aux règles de la prise de vue l'autorisent.

Lorsque l'utilisateur demande, à travers l'interface 30 une prise de son sur l'ensemble des personnes, le dispositif 20 commande la prise de son de manière à capter le champ sonore des différents participants. Ce champ sonore peut être obtenu de  
5 différentes manières :

1) A partir d'un champ large réalisé par l'antenne acoustique 12 ;

2) Par sommation des champs sonores issus de  
10 différents dispositifs de prise de son (microphones ou antenne acoustique), dispositifs sélectionnés à partir de l'analyse de la scène.

Lorsque l'utilisateur demande, à travers l'interface 30, une prise de son sur une personne, le  
15 dispositif 20 commande la prise de son de manière à capter le champ sonore des différentes personnes. Ce champ sonore peut être obtenu de différentes manières :

1) A partir du champ étroit et orientable de l'antenne acoustique 12. C'est le mode de réalisation  
20 réalisé décrit précédemment ;

2) A partir d'un ou plusieurs dispositifs de prise de son, dispositifs sélectionnés à partir des informations visuelles ou des informations audiovisuelles.

L'interface 30 permet à l'utilisateur du système  
25 d'obtenir une prise de vue et de son conformes à sa demande (un plan large sur l'ensemble des personnes, un plan serré sur une personne particulière). L'envoi d'une commande issue de l'interface, déclenche la  
30 commande d'orientation des capteurs de prise de son et de vue, en fonction de la scène audiovisuelle, analysée par le dispositif d'analyse de scène.

A cette fin, l'interface comporte une interface logique 31 et une interface physique 32.

L'interface physique 32 peut être réalisée selon différents modes de réalisation décrits dans la suite à propos des figures 3 à 7.

5 L'interface logique 31 est, selon un mode de réalisation préférentiel, constituée par un programme chargé dans le système de traitement du signal d'analyse de scène 40. Cette interface logique 31 récupère des informations de position des personnes sur la scène issues de traitement d'analyse de scène et  
10 établit une correspondance entre ces informations de position et les informations de sélection données à travers l'interface physique par l'intervenant.

Cette interface logique 31 interprète (c'est à dire qu'elle décode) les informations reçues de l'unité 40  
15 pour fournir des signaux de commande de position interprétables par le dispositif de commande 20 afin d'effectuer le cadrage désiré autour de la personne sélectionnée ou du groupe.

Plusieurs modes de réalisation de l'interface  
20 physique 32 sont donc donnés dans la suite.

Un premier mode de réalisation comprend une interface graphique 32A installée sur un micro-ordinateur ou station de travail P comme représenté sur la figure 3.

25 Avec une souris 320, l'utilisateur choisit d'obtenir une prise de vue et de son sur l'ensemble des personnes de la scène, en cliquant sur une fenêtre nommée "Ensemble", référencé E. Avec une souris, l'utilisateur choisit d'obtenir une prise de vue et de  
30 son sur une personne de la scène, en cliquant sur une fenêtre portant le numéro de la personne désirée P1-Pn ou du groupe de personnes.

Sur cette interface graphique 32A le libellé en chiffre des personnes peut être remplacé par l'image de



la personne 321 obtenue par le système d'analyse. Cette image est obtenue soit à un moment fixé par l'utilisateur du système, soit elle est automatiquement rafraîchie au cours de la réunion.

5 Une interface graphique 32A avec l'image des personnes 321 est plus ergonomique pour l'utilisateur, car l'interface affiche les prises de vues que peut sélectionner l'utilisateur.

10 On peut également prévoir un affichage par incrustation sur l'écran de visioconférence, en format réduit, de l'image des personnes présentes avec des numéros, à la visioconférence et, la sélection alors au moyen d'une télécommande.

15 Dans un autre mode de réalisation, la souris 320 peut être remplacée par un écran tactile et/ou par un dispositif de reconnaissance de la parole R.

20 Un autre mode de réalisation réalisé pour l'interface physique 32 est représenté par la figure 4. Selon ce mode de réalisation, on détourne l'usage de la télécommande 32B de la caméra d'analyse de la scène visuelle 60 pour permettre à l'utilisateur du système d'envoyer des signaux de commande à la caméra 60. Le détournement et l'utilisation de cette télécommande a été réalisé pour des raisons de facilité et de rapidité  
25 de mise en oeuvre.

La télécommande infrarouge 32B est en communication (commandes CDE) avec la caméra d'analyse 60. Cette télécommande de caméra d'analyse dispose d'un certain nombre de touches dont notamment des touches  
30 correspondant à des mémoires de position et une touche "home" H correspondant à la position de repos de la caméra.

Les mémoires de position ne sont pas utilisées en tant que telles pour pointer des directions de

l'espace, mais on n'utilise que le fait que les touches sont activées.

Les positions des mémoires de position sont préalablement initialisées par le système, à la position de repos de la caméra. La caméra d'analyse  
5 étant fixe dans un des modes de réalisation, le déclenchement des positions 1 à 6 ou de la touche "home" H n'a pas d'effet sur la position de cette caméra d'analyse 60.

10 Ainsi, en appuyant par exemple sur la touche "home" H, l'utilisateur déclenche via les dispositifs 60, 40, 30 et 20, une prise de vue et de son sur l'ensemble des personnes présentes dans la scène.

Et, en appuyant sur l'une des touches de 1 à 6 correspondant à la mémoire des positions, l'utilisateur déclenche via les dispositifs 60, 40, 30, et 20, une prise de vue sur la personne correspondante (6  
15 personnes maximum dans cette version).

D'autres télécommandes peuvent être utilisées qu'elles utilisent ou pas la caméra d'analyse en tant  
20 que relais de transmission.

On peut par exemple utiliser une télécommande à programmation universelle à la place de la télécommande de la caméra d'analyse, pour permettre à l'utilisateur  
25 de sélectionner la prise de vue et de son. Ce mode de réalisation n'est pas illustré car il correspond au schéma de la figure 4 à ceci près que la télécommande 32B est dans ce cas une télécommande universelle.

Dans une phase préparatoire du matériel en vue de  
30 mettre en oeuvre l'application, on aura au préalable enregistré dans la télécommande universelle certains des codes de la caméra d'analyse (mémoire de position, "home", ...). Lors de l'utilisation effective de la télécommande 32B, les signaux infrarouges envoyés par

la télécommande sont décodés et analysés par l'interface logique 31.

L'utilisation d'une télécommande universelle, permet d'une part, de rendre indisponibles certaines commandes de la caméra d'analyse 60 (la télécommande universelle n'apprend pas les codes correspondants), et d'autre part, d'avoir des touches dont le libellé correspond mieux à la demande de l'utilisateur : Ainsi, des touches "+" et "-" permettent à l'utilisateur de zapper d'une personne à une autre, comme il a l'habitude de le faire avec les chaînes de télévision.

On va se reporter maintenant au schéma de la figure 5 qui correspond à un autre mode de réalisation conforme à l'invention.

Afin de ne pas dépendre de la caméra d'analyse en tant que relais entre la télécommande et l'interface 31 et pour offrir à l'utilisateur une palette de commande plus riche, on peut utiliser un dispositif émetteur récepteur 70 du commerce.

Cet émetteur-récepteur 70 capte des signaux infrarouges CDE issus de la télécommande 32B et renvoie des codes vers l'interface logique 31, par exemple à travers un port de communication RS232, connecté à l'interface 30.

La figure 6 illustre un mode de réalisation selon lequel l'interface physique 32 comporte une commande à distance par la voix 32B associée à un dispositif de reconnaissance de la parole 80 existant.

La figure 7 illustre un mode de réalisation selon lequel l'interface physique 32 comporte une commande à distance par le geste 32B associée à un dispositif de reconnaissance du geste 90 existant.

Les précédents modes de réalisation de l'interface permettent à un utilisateur de sélectionner les prises

de vue et de son obtenues avec les capteurs physiquement présents dans la salle où se trouve l'utilisateur. Cela est symbolisé par le signal A (local) servant à l'analyse de la scène sur la figure

5 1.

Dans un autre mode de réalisation, les interfaces 31, 32 précédemment décrites permettent de commander les capteurs de prise de vue et de son physiquement présents dans une salle distante ( où ne se trouve pas l'utilisateur), la salle avec laquelle il est en visioconférence par exemple.

Ainsi, l'utilisateur participant à une visioconférence, sélectionne et obtient les prises de vue et de son désirées. Pour cela, le signal A' (distant) pour l'analyse de scène ou correspondant à l'analyse (signal déjà analysé par un système distant dépourvu du signal vidéo) sera appliqué à une entrée du dispositif d'analyse 40.

Dans ce mode de réalisation, les signaux C' émis par la télécommande infrarouge ou par l'interface graphique (et les signaux nécessaires à la constitution des images de l'interface graphique) sont transportés avec l'image, le son et les autres signaux de la visioconférence.

25 Dans ce mode de réalisation, l'éventuel conflit de commande des capteurs entre la salle locale et la salle distante doit être géré.

## REVENDICATIONS

1. Système automatique de prise de son et d'images notamment pour visioconférence, comportant des moyens de commande (20) des capteurs de prise de vues et de son (10), des moyens d'analyse de scène (40) pour  
5 fournir des signaux de position aux moyens de commande (20), des moyens de sélection (30) d'une personne ou d'un groupe parmi des personnes de la scène filmée,

- caractérisé en ce que les moyens de sélection comportent une interface physique (32) comprenant une  
10 commande à distance (32A, 32B) apte à permettre la sélection de l'une quelconque des personnes de la scène ou d'un groupe, pour avoir un cadrage automatique autour de cette personne ou du groupe, ou de sélectionner l'ensemble des personnes pour avoir un  
15 cadrage général de la scène ;

- et en ce que les moyens de cadrage comportent une interface logique (31) apte à établir une correspondance entre la personne sélectionnée par la  
20 commande à distance (32A ou 32B) et les informations de position issues de l'analyse de scène pour fournir aux moyens de commande (20) les informations de position de cette personne ou du groupe par rapport à la scène filmée.

25 2. Système selon la revendication 1, caractérisé en ce que la commande à distance (32B) est une télécommande universelle, activant un dispositif apte à émettre des signaux de commande à l'interface logique (31) .

30 3. Système selon la revendication 2, caractérisé en ce que les signaux de commande de ladite télécommande

(32B) sont reçus et ré-émis par un émetteur-récepteur (70).

5        4. Système selon la revendication 2, caractérisé en ce que les signaux de commande de ladite télécommande (32B) sont reçus et ré-émis par un dispositif de reconnaissance de la parole (80) ou de reconnaissance de geste (90).

10       5. Système selon la revendication 1, caractérisé en ce que la commande à distance (32B) est la télécommande de la caméra d'analyse d'image (60), les signaux de commande de ladite télécommande (32) étant reçus et ré-émis par la caméra d'analyse (60) vers l'interface  
15       logique (31).

20       6. Système selon la revendication 1, caractérisé en ce que commande à distance (32B) est une télécommande universelle, les signaux de commande de ladite télécommande (32B) étant reçus et ré-émis par la caméra d'analyse (60).

25       7. Système selon la revendication 1, caractérisé en ce que la commande à distance comprend une interface graphique (32A).

30       8. Système de visioconférence selon la revendication 7, caractérisé en ce que la commande à distance comporte en outre un écran (321) sur lequel sont visualisées la scène et les différentes zones sélectionnables.

9. Système selon la revendication 7, caractérisé en ce que la commande à distance comporte un périphérique

(320) d'entrée/sortie d'ordinateur pour sélectionner les zones repérées.

5           10. Système de visioconférence selon l'une  
quelconque des revendications précédentes, selon lequel  
les moyens d'analyse de scène reçoivent un signal  
d'analyse local (A), caractérisé en ce que les moyens  
de sélection permettent de sélectionner une personne ou  
un groupe de personnes de la scène filmée localement et  
10   en ce que, les moyens de cadrage automatique utilisent  
les informations de la scène filmée localement.

          11. Système de visioconférence selon l'une  
quelconque des revendications précédentes recevant un  
15   signal (A') d'un système distant pour ou correspondant  
à l'analyse de scène, caractérisé en ce que les moyens  
de sélection permettent de sélectionner une personne ou  
un groupe de personnes de la scène filmée à distance et  
en ce que les moyens de cadrage automatique permettent  
20   de piloter le cadrage de la scène filmée à distance,  
les signaux de commande étant transportés vers le  
système distant.

FIG. 1

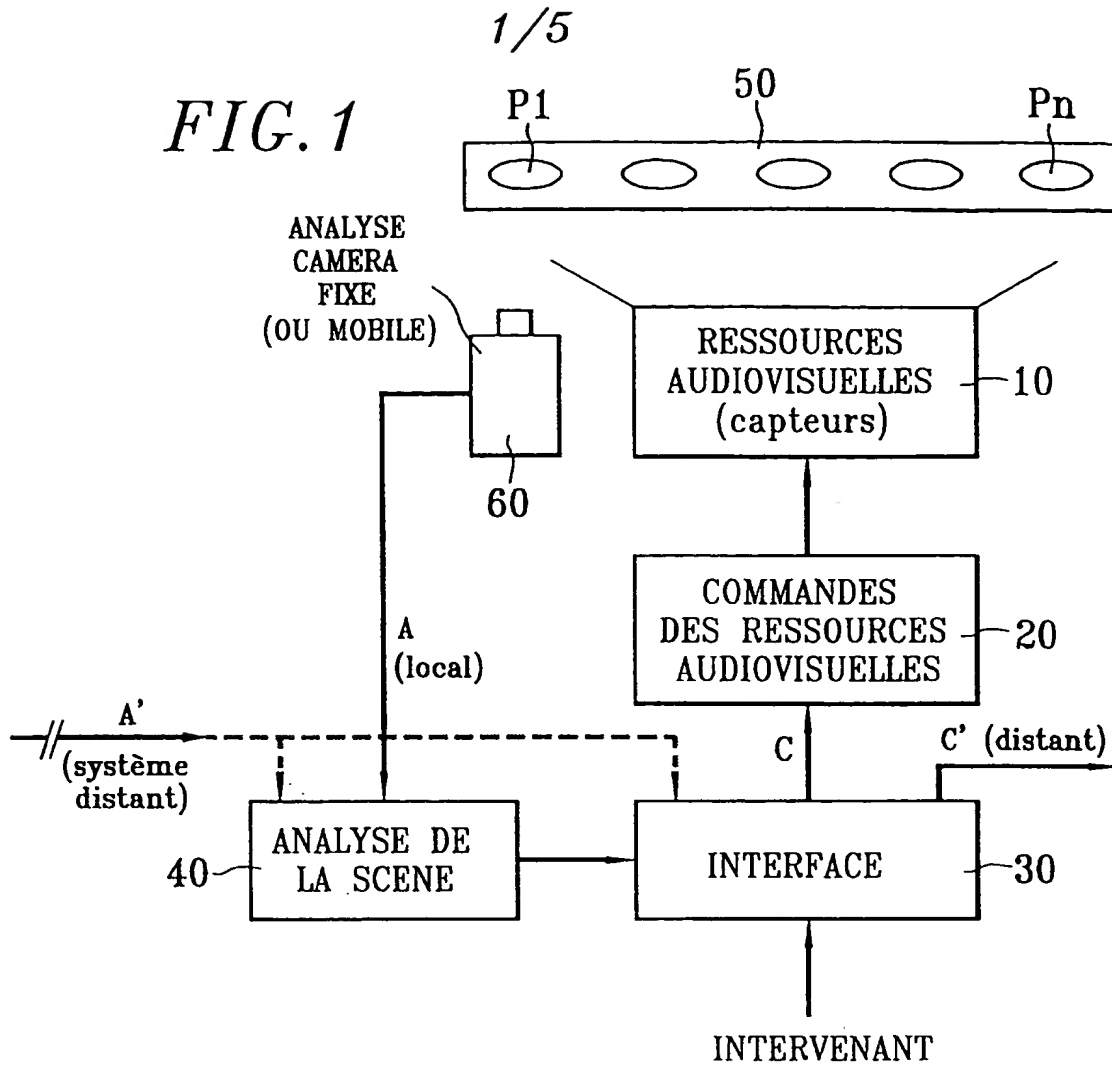
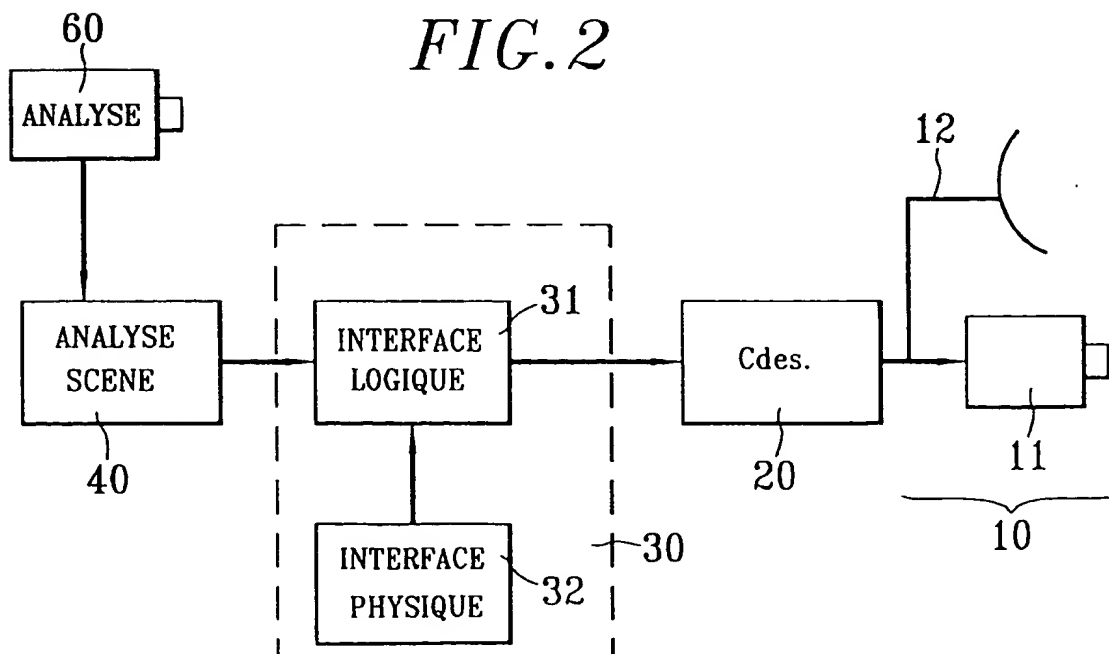


FIG. 2





2/5

FIG. 3

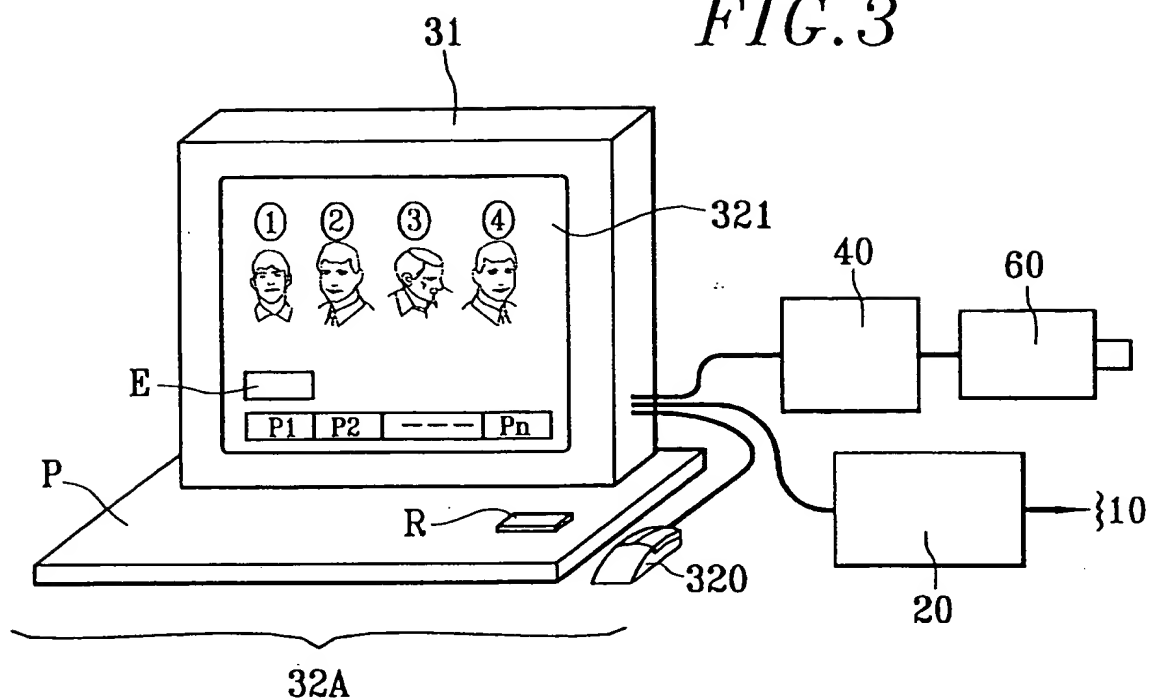
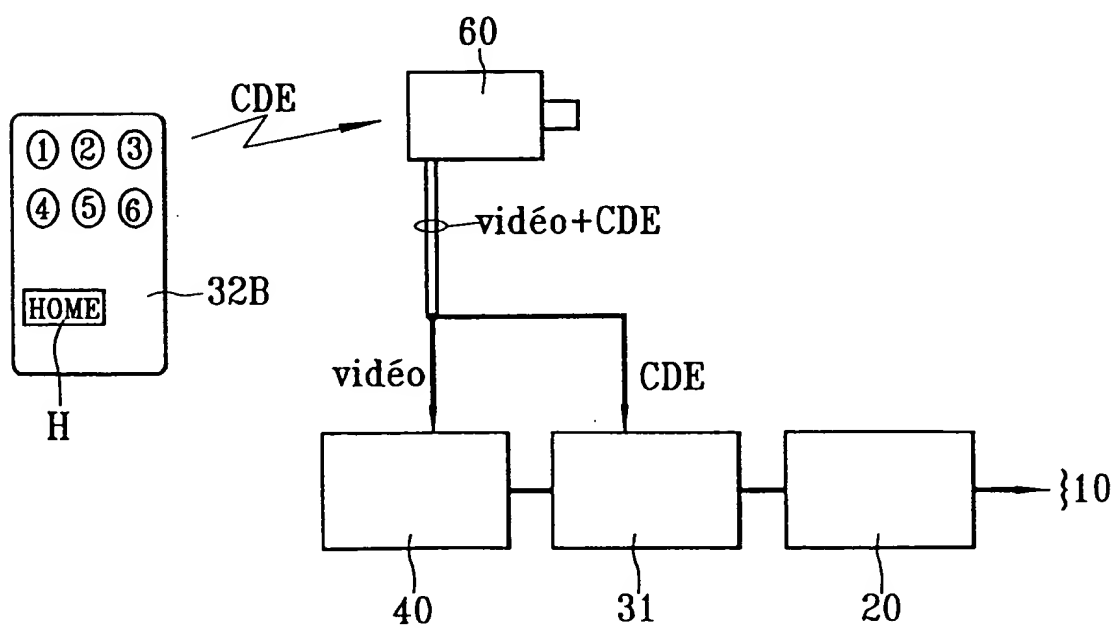
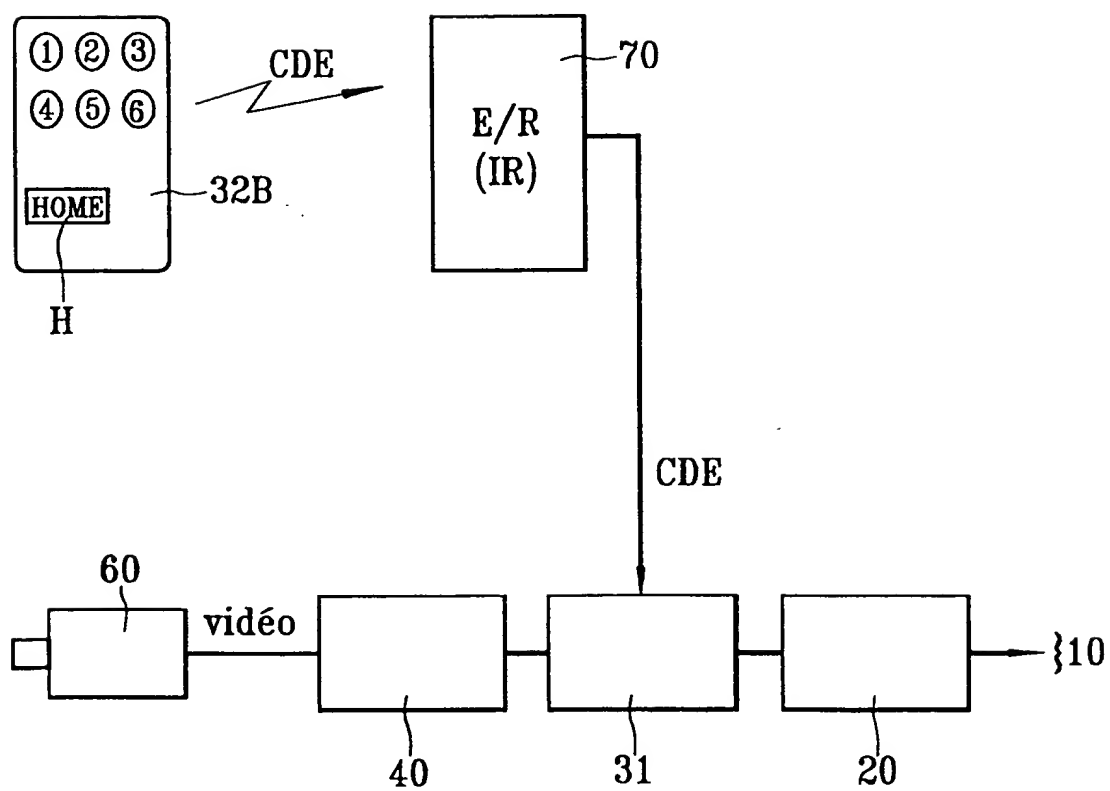


FIG. 4



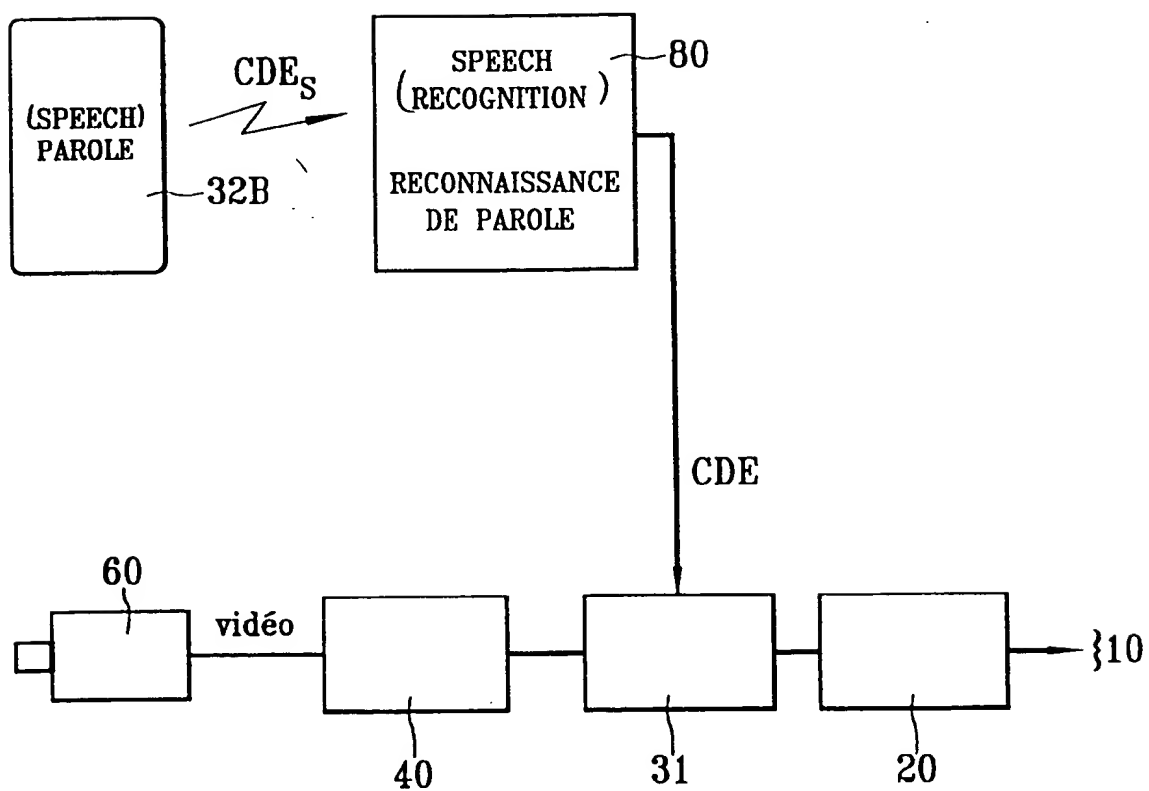
3/5

FIG. 5



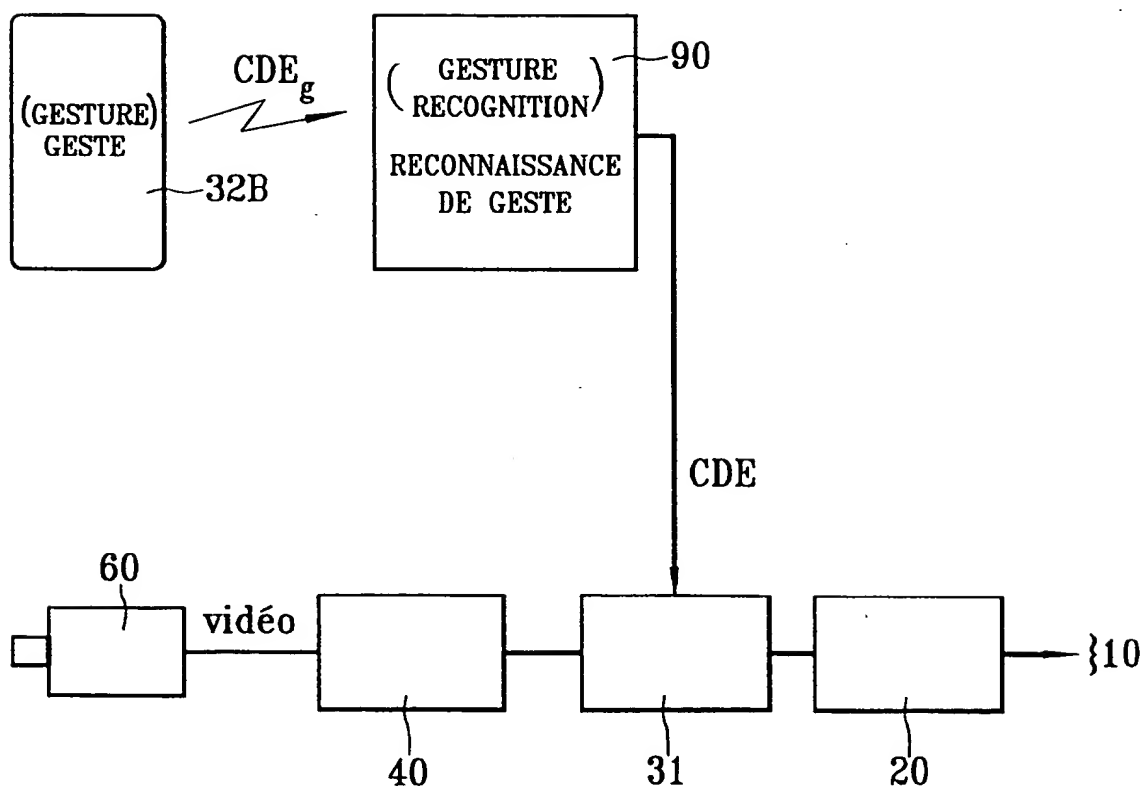
4/5

FIG. 6



5/5

FIG. 7



# INTERNATIONAL SEARCH REPORT

Int. National Application No

PCT/FR 99/02047

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 H04N7/15 H04N7/14

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04N 601S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 14587 A (DEVITECH IS ;DUE JESPER (DK); HEDER THOMAS (DK); JUERGENSEN PETER) 17 May 1996 (1996-05-17)	1,2,7,9, 10
Y	page 1, line 8 - line 12 page 2, line 9 - line 14 page 3, line 16 - line 21 page 5, line 3 - line 14 page 8, line 24 -page 9, line 3 claim 7	8
Y	GB 2 252 473 A (RADAMEC EPO LIMITED) 5 August 1992 (1992-08-05)	8
A	abstract page 1, line 21 -page 2, line 18 page 4, line 12 - line 20 -/-	1,2,7

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

25 November 1999

Date of mailing of the international search report

03/12/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentstrasse 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Marie-Julie, J-M

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 99/02047

## C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 434 617 A (BIANCHI MICHAEL H) 18 July 1995 (1995-07-18) abstract column 1, line 59 -column 2, line 6 column 4, line 4 - line 8	1,2,10, 11
A	WO 95 11566 A (VIDEO CONFERENCING SYS INC) 27 April 1995 (1995-04-27) page 5, line 22 -page 6, line 24 page 15, line 32 -page 16, line 6 page 21, line 9 -page 26, line 35	1-3,5,6, 10,11
A	US 5 686 957 A (BAKER ROBERT G) 11 November 1997 (1997-11-11) abstract column 1, line 45 - line 65 column 5, line 48 - line 54 column 6, line 37 - line 42	1,2,10, 11
A	US 5 745 161 A (ITO KAN) 28 April 1998 (1998-04-28) abstract column 1, line 56 -column 2, line 24	1,2,7-9
A	US 4 274 609 A (FERRIER NOEL H L ET AL) 23 June 1981 (1981-06-23) abstract	1,2,7-9
A	EP 0 751 473 A (LUCENT TECHNOLOGIES INC) 2 January 1997 (1997-01-02)	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Serial Application No

PCT/FR 99/02047

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9614587 A	17-05-1996	AT 176951 T AU 3802395 A DE 69507963 D DE 69507963 T EP 0789846 A ES 2131338 T JP 10508713 T	15-03-1999 31-05-1996 01-04-1999 07-10-1999 20-08-1997 16-07-1999 25-08-1998
GB 2252473 A	05-08-1992	WO 9306690 A	01-04-1993
US 5434617 A	18-07-1995	CA 2148231 A,C WO 9417636 A	04-08-1994 04-08-1994
WO 9511566 A	27-04-1995	AU 7921094 A CA 2174336 A CN 1135823 A EP 0724809 A JP 9506217 T US 5528289 A US 5589878 A US 5515099 A US 5583565 A US 5598209 A US 5526037 A US 5568183 A	08-05-1995 27-04-1995 13-11-1996 07-08-1996 17-06-1997 18-06-1996 31-12-1996 07-05-1996 10-12-1996 28-01-1997 11-06-1996 22-10-1996
US 5686957 A	11-11-1997	US 5508734 A BR 9502919 A CA 2152314 A EP 0695085 A JP 8055215 A	16-04-1996 05-03-1996 28-01-1996 31-01-1996 27-02-1996
US 5745161 A	28-04-1998	JP 7123390 A	12-05-1995
US 4274609 A	23-06-1981	FR 2389865 A DE 2819938 A GB 1577108 A	01-12-1978 16-11-1978 15-10-1980
EP 0751473 A	02-01-1997	US 5805745 A CA 2177639 A JP 9102043 A	08-09-1998 27-12-1996 15-04-1997

PCT/FR 99/02047

Formulate PCT/SA21D (daughter file) 0 April 1993



# RAPPORT DE RECHERCHE INTERNATIONALE

Document International No

PCT/FR 99/02047

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	US 5 434 617 A (BIANCHI MICHAEL H) 18 juillet 1995 (1995-07-18) abrégé colonne 1, ligne 59 - colonne 2, ligne 6 colonne 4, ligne 4 - ligne 8	1,2,10, 11
A	WO 95 11566 A (VIDEO CONFERENCING SYS INC) 27 avril 1995 (1995-04-27) page 5, ligne 22 - page 6, ligne 24 page 15, ligne 32 - page 16, ligne 6 page 21, ligne 9 - page 26, ligne 35	1-3,5,6, 10,11
A	US 5 686 957 A (BAKER ROBERT G) 11 novembre 1997 (1997-11-11) abrégé colonne 1, ligne 45 - ligne 65 colonne 5, ligne 48 - ligne 54 colonne 6, ligne 37 - ligne 42	1,2,10, 11
A	US 5 745 161 A (ITO KAN) 28 avril 1998 (1998-04-28) abrégé colonne 1, ligne 56 - colonne 2, ligne 24	1,2,7-9
A	US 4 274 609 A (FERRIER NOEL H L ET AL) 23 juin 1981 (1981-06-23) abrégé	1,2,7-9
A	EP 0 751 473 A (LUCENT TECHNOLOGIES INC) 2 janvier 1997 (1997-01-02)	

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

De l'Office International No

PCT/FR 99/02047

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9614587 A	17-05-1996	AT 176951 T AU 3802395 A DE 69507963 D DE 69507963 T EP 0789846 A ES 2131338 T JP 10508713 T	15-03-1999 31-05-1996 01-04-1999 07-10-1999 20-08-1997 16-07-1999 25-08-1998
GB 2252473 A	05-08-1992	WO 9306690 A	01-04-1993
US 5434617 A	18-07-1995	CA 2148231 A,C WO 9417636 A	04-08-1994 04-08-1994
WO 9511566 A	27-04-1995	AU 7921094 A CA 2174336 A CN 1135823 A EP 0724809 A JP 9506217 T US 5528289 A US 5589878 A US 5515099 A US 5583565 A US 5598209 A US 5526037 A US 5568183 A	08-05-1995 27-04-1995 13-11-1996 07-08-1996 17-06-1997 18-06-1996 31-12-1996 07-05-1996 10-12-1996 28-01-1997 11-06-1996 22-10-1996
US 5686957 A	11-11-1997	US 5508734 A BR 9502919 A CA 2152314 A EP 0695085 A JP 8055215 A	16-04-1996 05-03-1996 28-01-1996 31-01-1996 27-02-1996
US 5745161 A	28-04-1998	JP 7123390 A	12-05-1995
US 4274609 A	23-06-1981	FR 2389865 A DE 2819938 A GB 1577108 A	01-12-1978 16-11-1978 15-10-1980
EP 0751473 A	02-01-1997	US 5805745 A CA 2177639 A JP 9102043 A	08-09-1998 27-12-1996 15-04-1997